

# **IP Protection for Generative AI**

## **- Foundation Models and IP -**

September 2023

Jung-Hyon (Jamie) Jun Esq.  
Lee & Jun Law Office

# What is Happening Now in AI?

- Rise of the Foundation Model

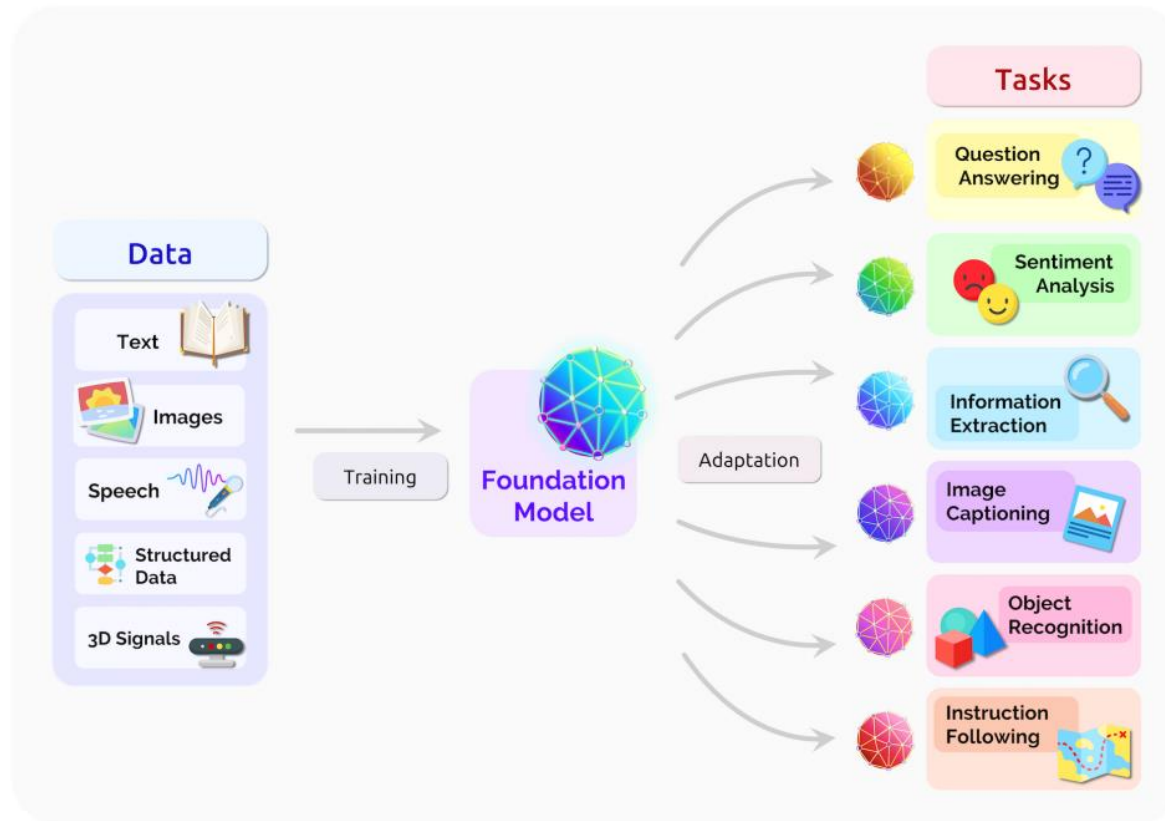
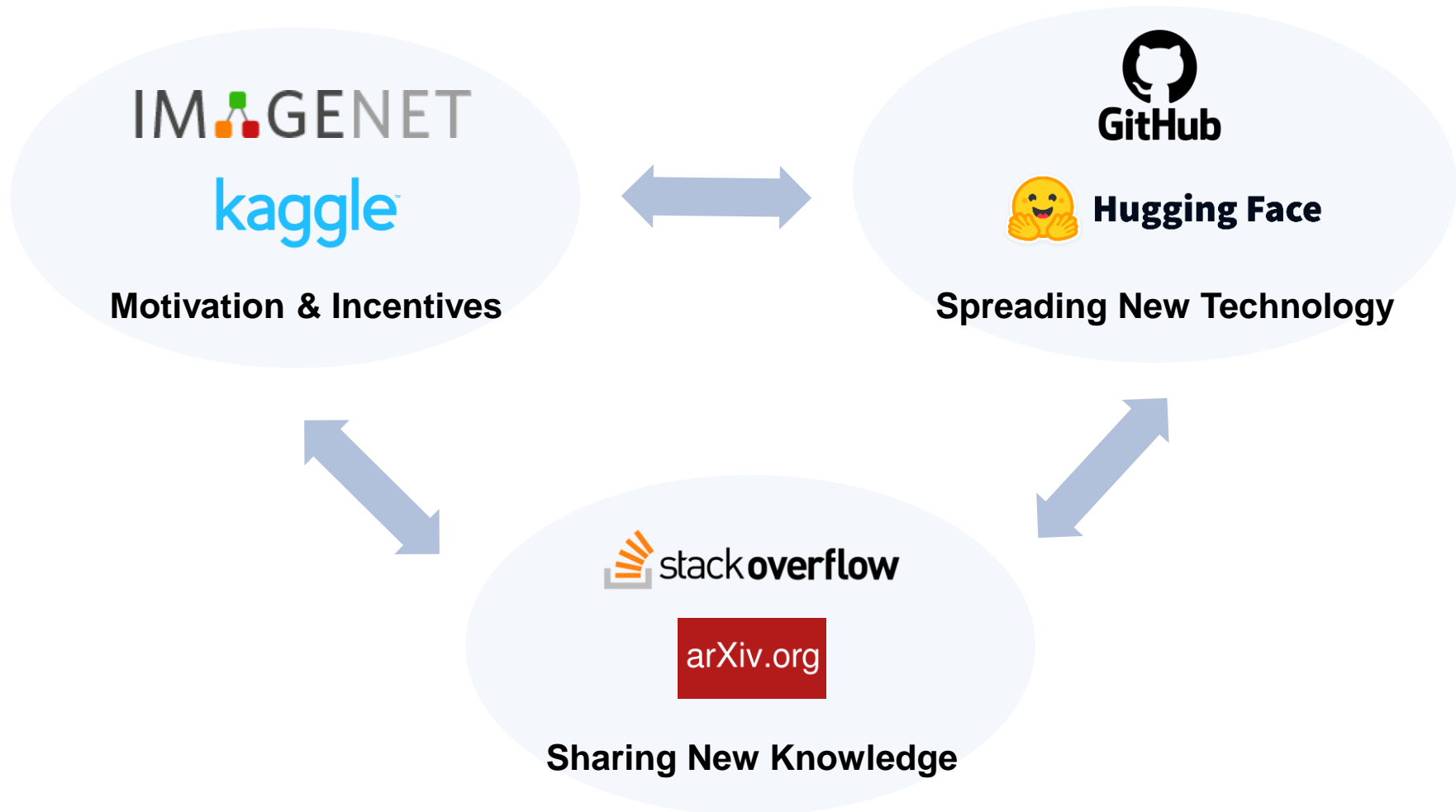


Fig. 2. A foundation model can centralize the information from all the data from various modalities. This one model can then be adapted to a wide range of downstream tasks.

Source: Bommasani et al. (2021)

# Collaborative Cycle in AI Development



# The End of Open Innovation?

---

## GPT-4 Technical Report

---

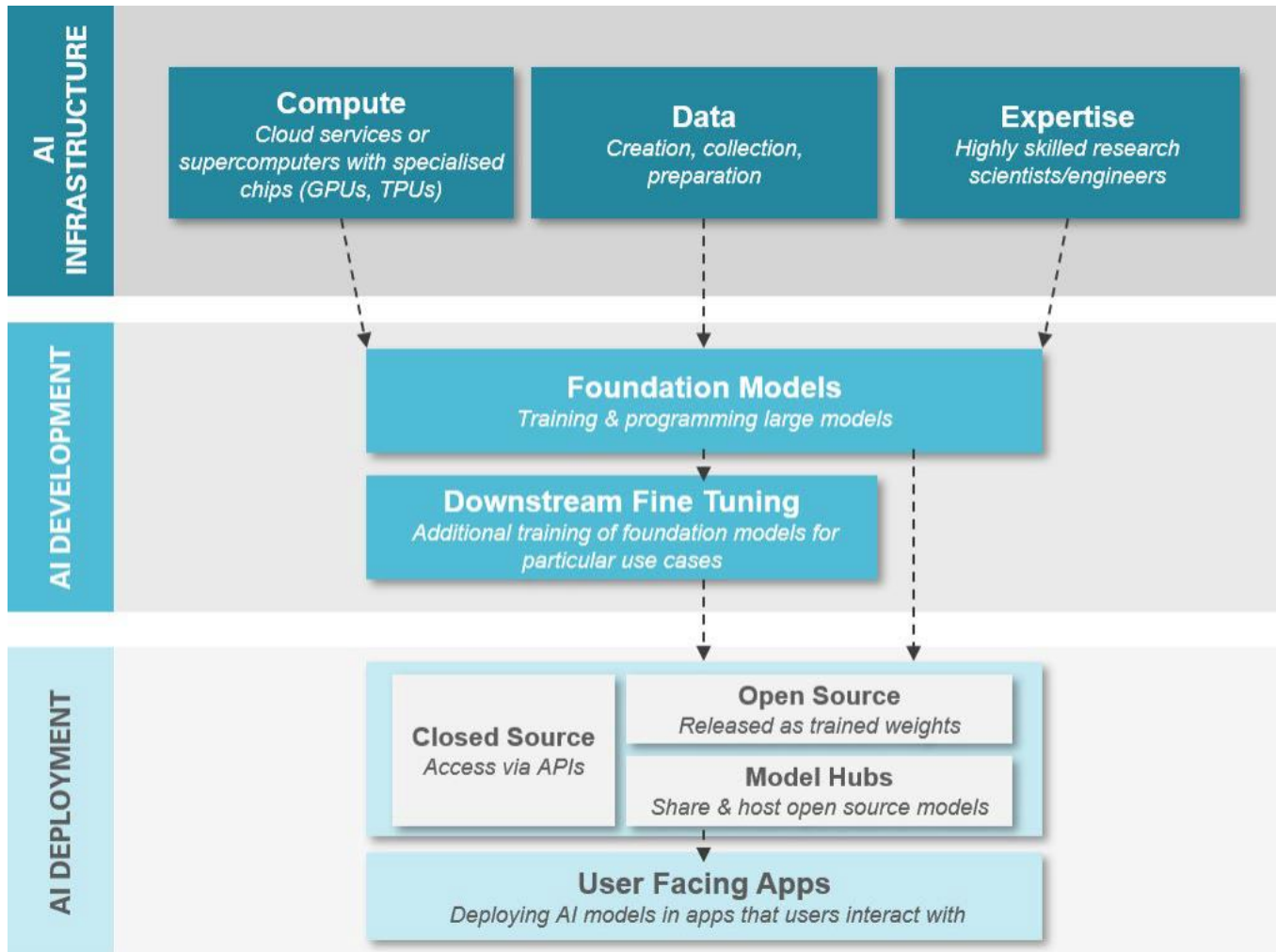
OpenAI\*

### 2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

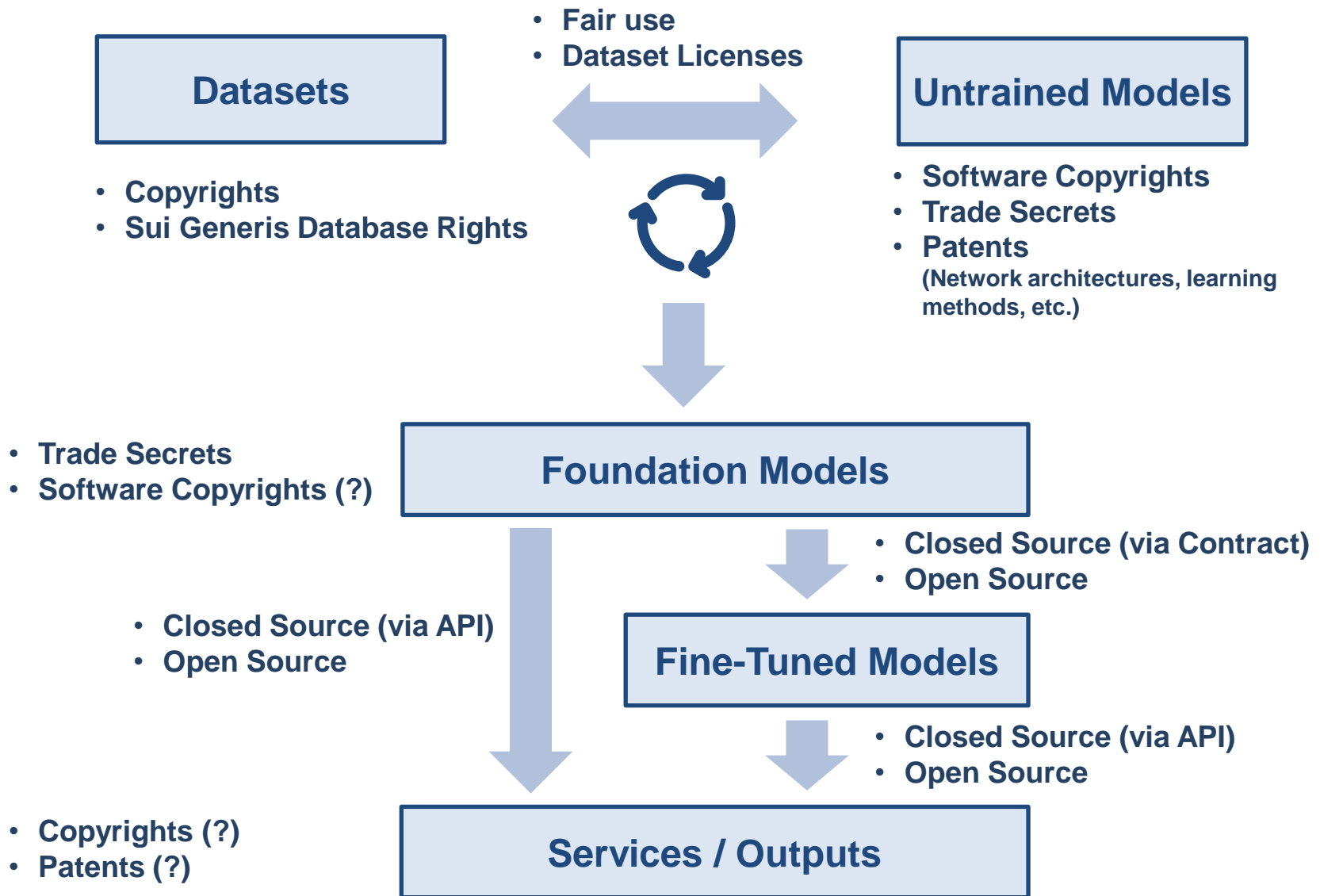
Source: OpenAI (2023)  
(Emphasis added.)

# Landscape of AI Ecosystem

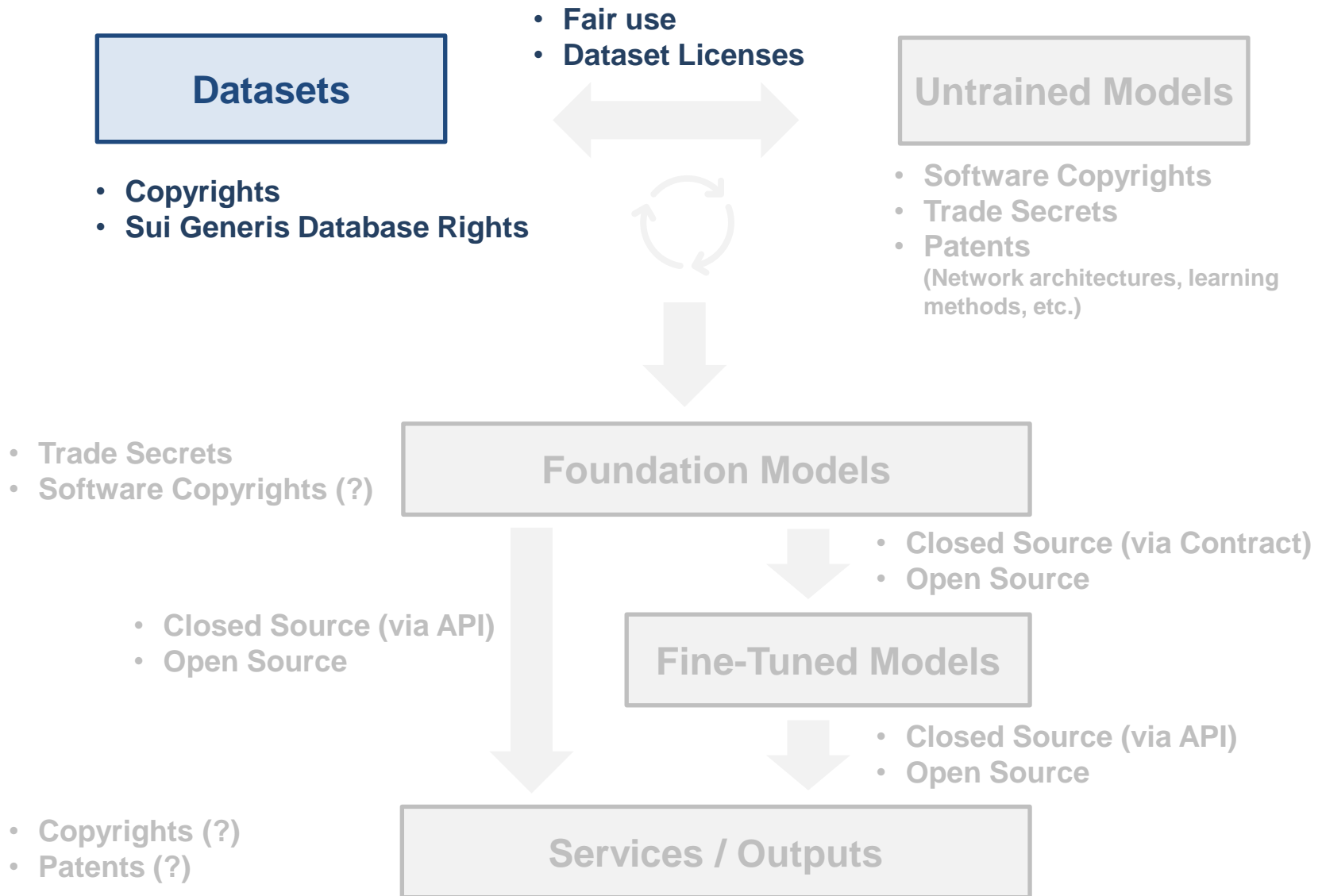


Source: Competition and Markets Authority (UK) (2023)  
(Emphasis added.)

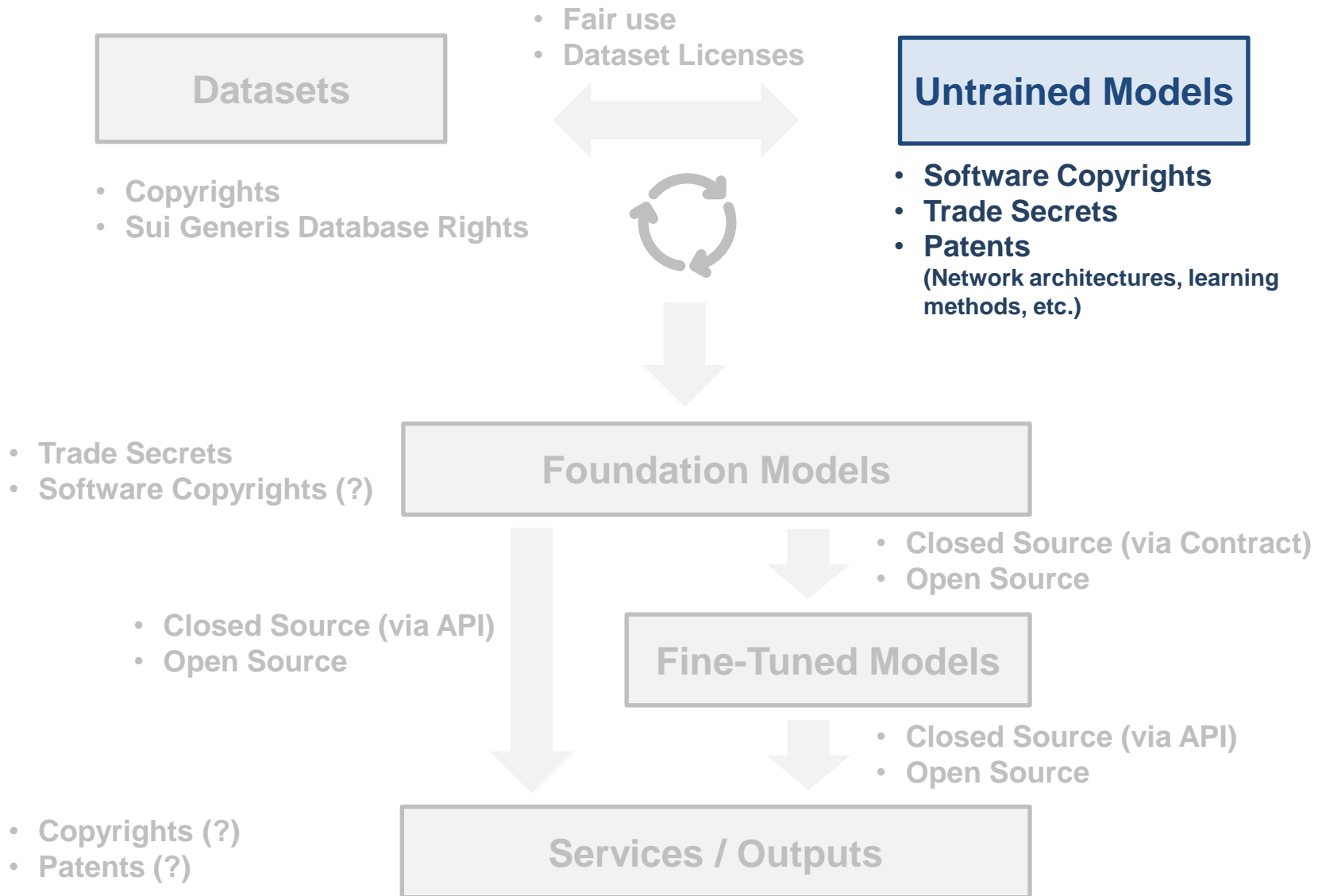
# Landscape of IP Issues



# IP Protection of Datasets



# IP Protection of AI Technologies





# Relaxed Grip on AI Patents

- Patents with CPC code G06N 3/08 (learning methods)

Google Patents (G06N3/08)

G06N3/08;

More than 100,000 results

Download Side-by-side

Sort by · Relevance · Group by · None · Deduplicate by · Family · Results / page · 10

**Generating synthetic data examples as interpolation of two data examples that ...**  
 WO EP US CN • US11687788B2 • James K. Baker • D5Ai Llc  
 Priority 2017-09-28 • Filed 2022-07-28 • Granted 2023-06-27 • Published 2023-06-27  
 Computer systems and methods generate data examples by training, through machine learning, a data generator with a training objective to produce a data example for a specific value of R, where R is value related to S1(x) and S2(x), where, for a data example, x, generated by the data generator, S1( ...

**Iterative training of a nodal network with data influence weights**  
 WO EP US CN • US11010671B2 • James K. Baker • D5Ai Llc  
 Priority 2018-01-30 • Filed 2020-07-15 • Granted 2021-05-18 • Published 2021-05-18

Assignees	Inventors	CPCs
Google Llc		3.5%
Intel Corporation		3.5%
Nvidia Corporation		3.3%
Tencent Technology (Shenzhen) Company Limited		2.6%
Samsung Electronics Co., Ltd.		2%

Source: Google Patents (Emphasis added.)

- Example

– Google’s patents on core technologies

FORBES > INNOVATION > AI

## Can Google Challenge OpenAI With Self-Attention Patents?

Alex Zhavoronkov, PhD Contributor @  
 Expert in AI for healthcare and longevity biotechnology

Follow

Jan 23, 2023, 06:29pm EST

### Google Holds Critical IP in Systems Utilizing Self-Attention

And when it comes to IP protection in AI, I don’t know any company stronger than Google. I remember our teams spending weeks ensuring that the architectures are unique enough to avoid the possible infringement of Google’s IP. Even DeepMind, a wholly-owned subsidiary of Alphabet, holds a very strong IP position in the space. A simple WIPO search returns over 800 results.

Source: Forbes

r/MachineLearning  
 by t0t0t4t4 2 mo. ago



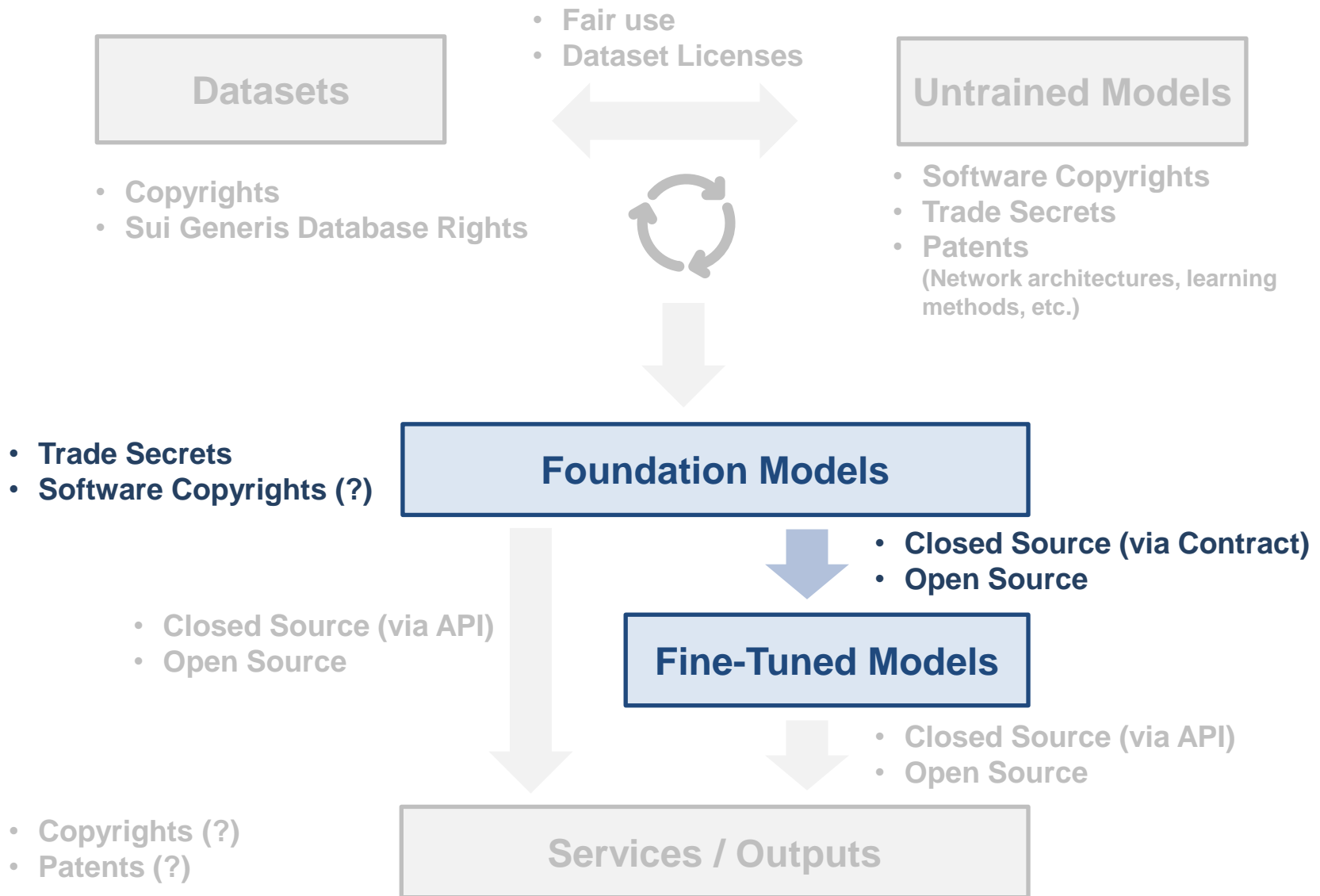
### [D] Can Google sue OpenAI for using the Transformer in their products?

As far as I know, the Transformer architecture is patented: <https://patents.google.com/patent/US10452978B2/en>. Since OpenAI has used the Transformer extensively (including GPT), I'm wondering if this can be considered as patent infringement.

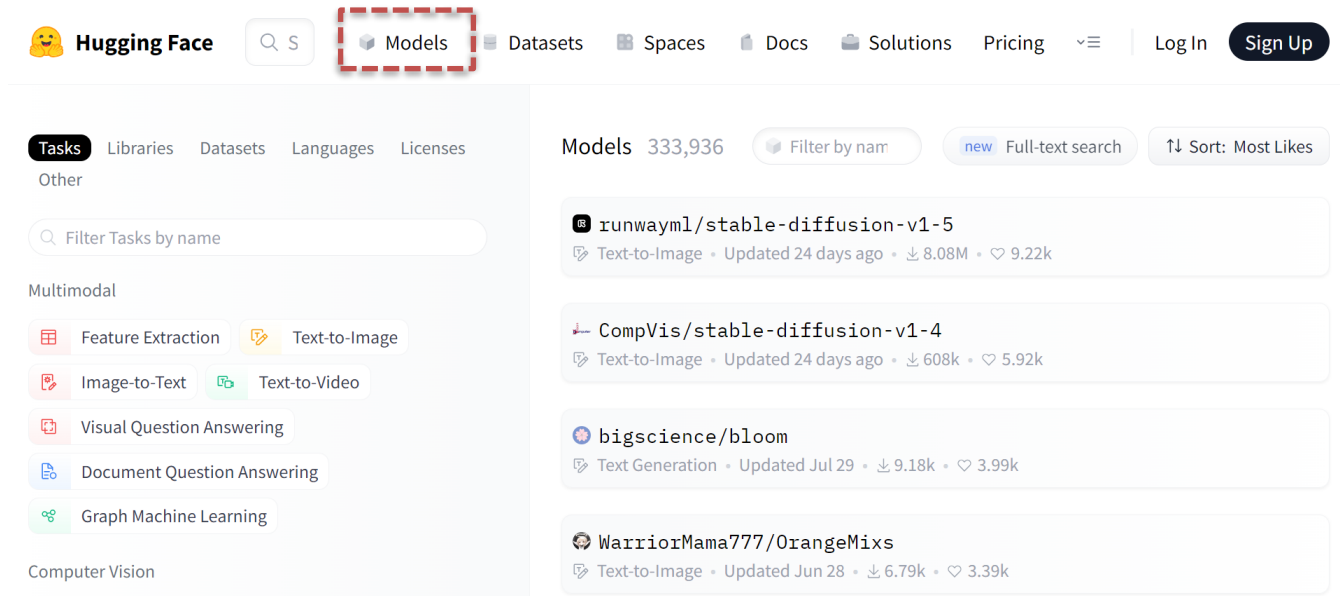
If you know about legal stuffs please share your opinions.

Source: Reddit

# IP Protection of Trained Models



# IP Protection of Trained Models



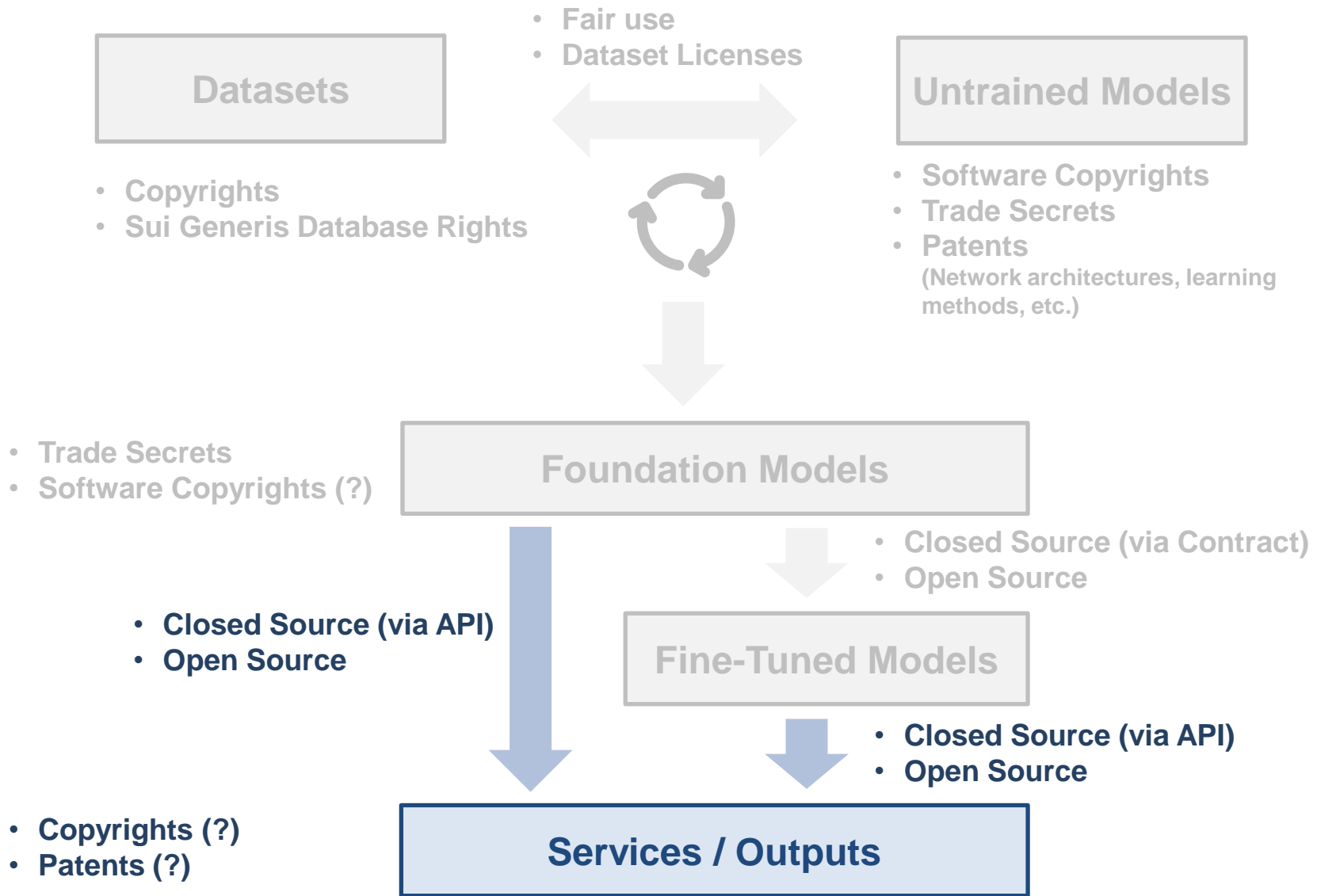
## 1. Download the weights

**IP Protection?**

- v1-5-pruned-emaonly.ckpt - 4.27GB, ema-only weight. uses less VRAM - suitable for inference
- v1-5-pruned.ckpt - 7.7GB, ema+non-ema weights. uses more VRAM - suitable for fine-tuning

Source: <https://huggingface.co/>  
(Emphasis added.)

# IP Protection of AI-Generated Outputs



# Open Foundation Models: A Game Changer?

Introducing Llama 2

The next generation of our open source large language model

Llama 2 is available for free for research and commercial use.

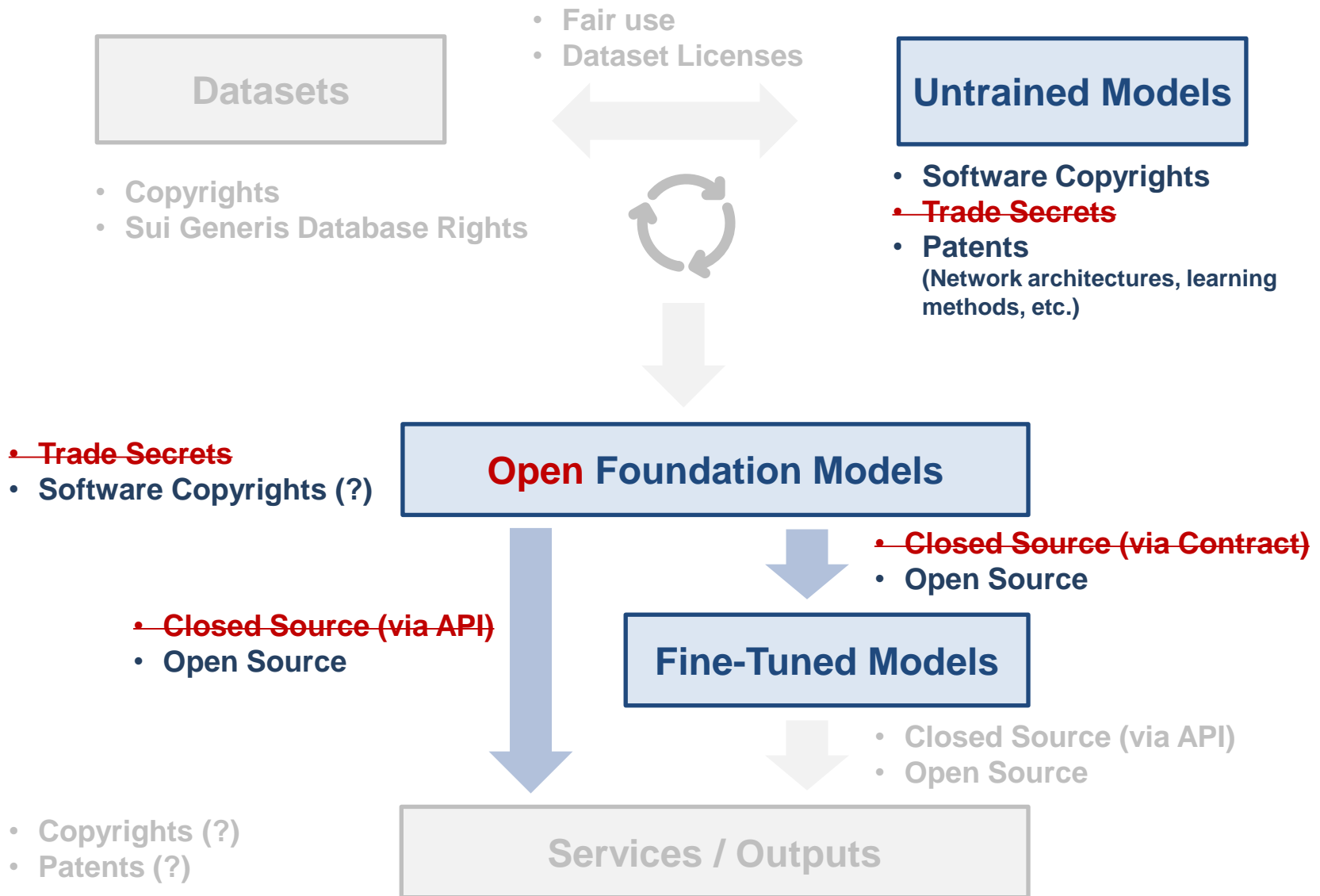
Download the Model

Benchmark (shots)	GPT-3.5	GPT-4	PaLM	PaLM-2-L	LLAMA 2
MMLU (5-shot)	70.0	<b>86.4</b>	69.3	78.3	68.9
TriviaQA (1-shot)	–	–	81.4	<b>86.1</b>	85.0
Natural Questions (1-shot)	–	–	29.3	<b>37.5</b>	33.0
GSM8K (8-shot)	57.1	<b>92.0</b>	56.5	80.7	56.8
HumanEval (0-shot)	48.1	<b>67.0</b>	26.2	–	29.9
BIG-Bench Hard (3-shot)	–	–	52.3	<b>65.7</b>	51.2

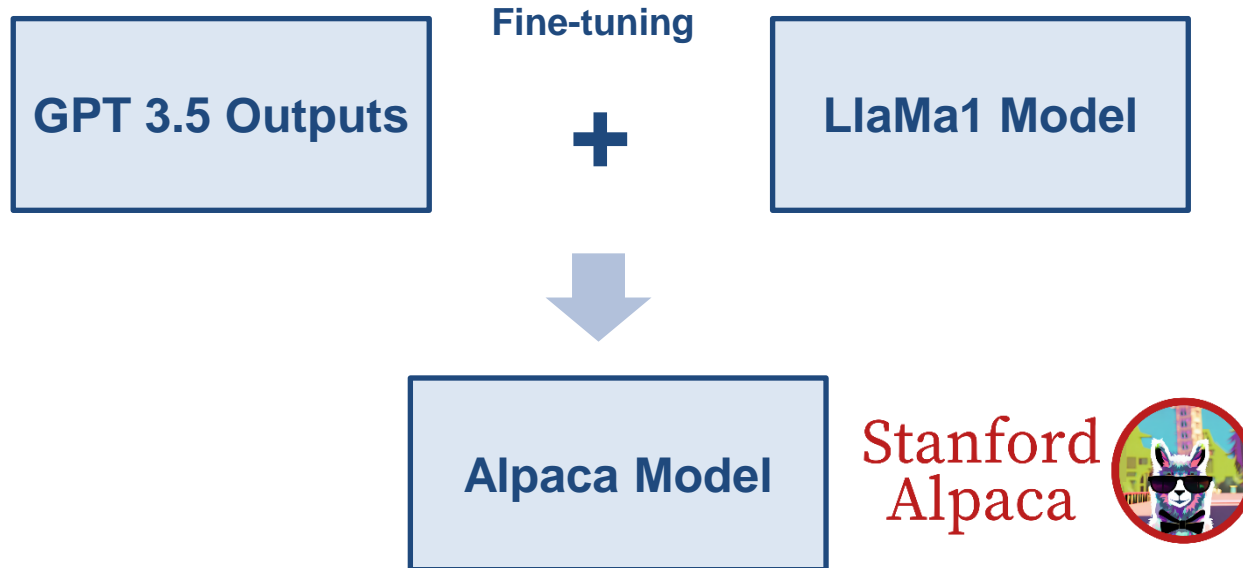
**Table 4: Comparison to closed-source models** on academic benchmarks. Results for GPT-3.5 and GPT-4 are from OpenAI (2023). Results for the PaLM model are from Chowdhery et al. (2022). Results for the PaLM-2-L are from Anil et al. (2023).

Source: <https://ai.meta.com/llama/>  
(Emphasis added.)

# IP Protection of **Open** Foundation Models



# Use Model Outputs to Train Another Model



- Training costs < \$600
- Performance comparable to GPT-3.5 (in certain tasks)

<https://crfm.stanford.edu/2023/03/13/alpaca.html>

# Use Model Outputs to Train Another Model

- **Potential violation of OpenAI's Terms of Use**
  - Legal question: Is this clause valid under applicable competition laws?

## 2. Usage Requirements

(a) **Use of Services.** You may access, and we grant you a non-exclusive right to use, the Services in accordance with these Terms. You will comply with these Terms and all applicable laws when using the Services. We and our affiliates own all rights, title, and interest in and to the Services.

(b) **Feedback.** We appreciate feedback, comments, ideas, proposals and suggestions for improvements. If you provide any of these things, we may use it without restriction or compensation to you.

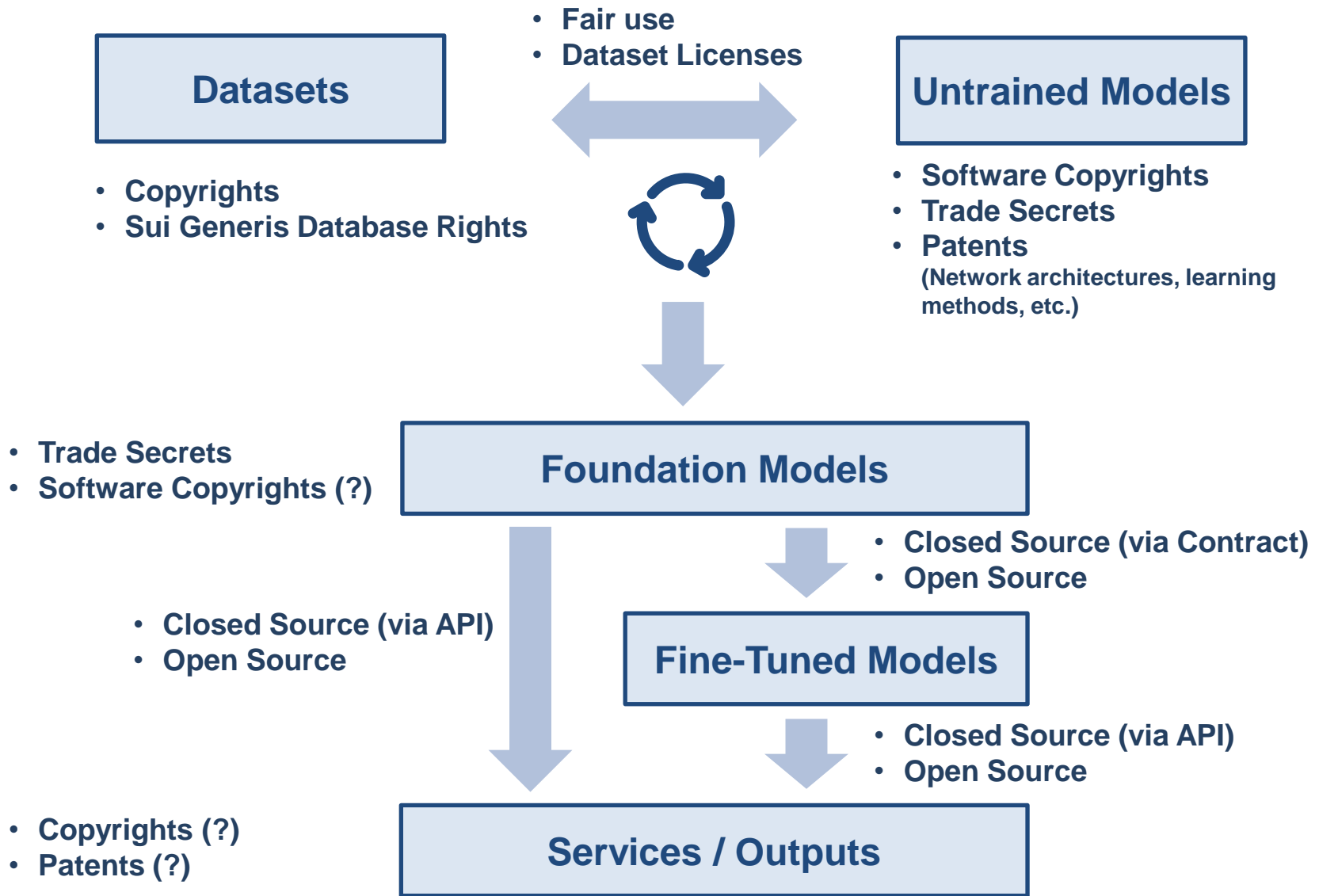
(c) **Restrictions.** You may not (i) use the Services in a way that infringes, misappropriates or violates any person's rights; (ii) reverse assemble, reverse compile, decompile, translate or otherwise attempt to discover the source code or underlying components of models, algorithms, and systems of the Services (except to the extent such restrictions are contrary to applicable law); (iii) use output from the Services to develop models that compete with OpenAI; (iv) except as permitted through the API, use any automated or programmatic method to extract data or output from the Services, including scraping, web harvesting, or web data extraction; (v) represent that output from the Services was human-generated when it is not or otherwise violate our Usage Policies; (vi) buy, sell, or transfer API keys without our prior consent; or (vii), send us any personal information of children under 13 or the applicable age of digital consent. You will comply with any rate limits and other requirements in our documentation. You may use Services only in geographies currently supported by OpenAI.

(c) **Restrictions.** You may not ...  
(iii) use output from the Services to develop models that compete with OpenAI;

Source: <https://openai.com/policies/terms-of-use>  
(Emphasis added.)



# Takeaway



**Thank You!**